

Knowledge Containers tagged to Taxonomies

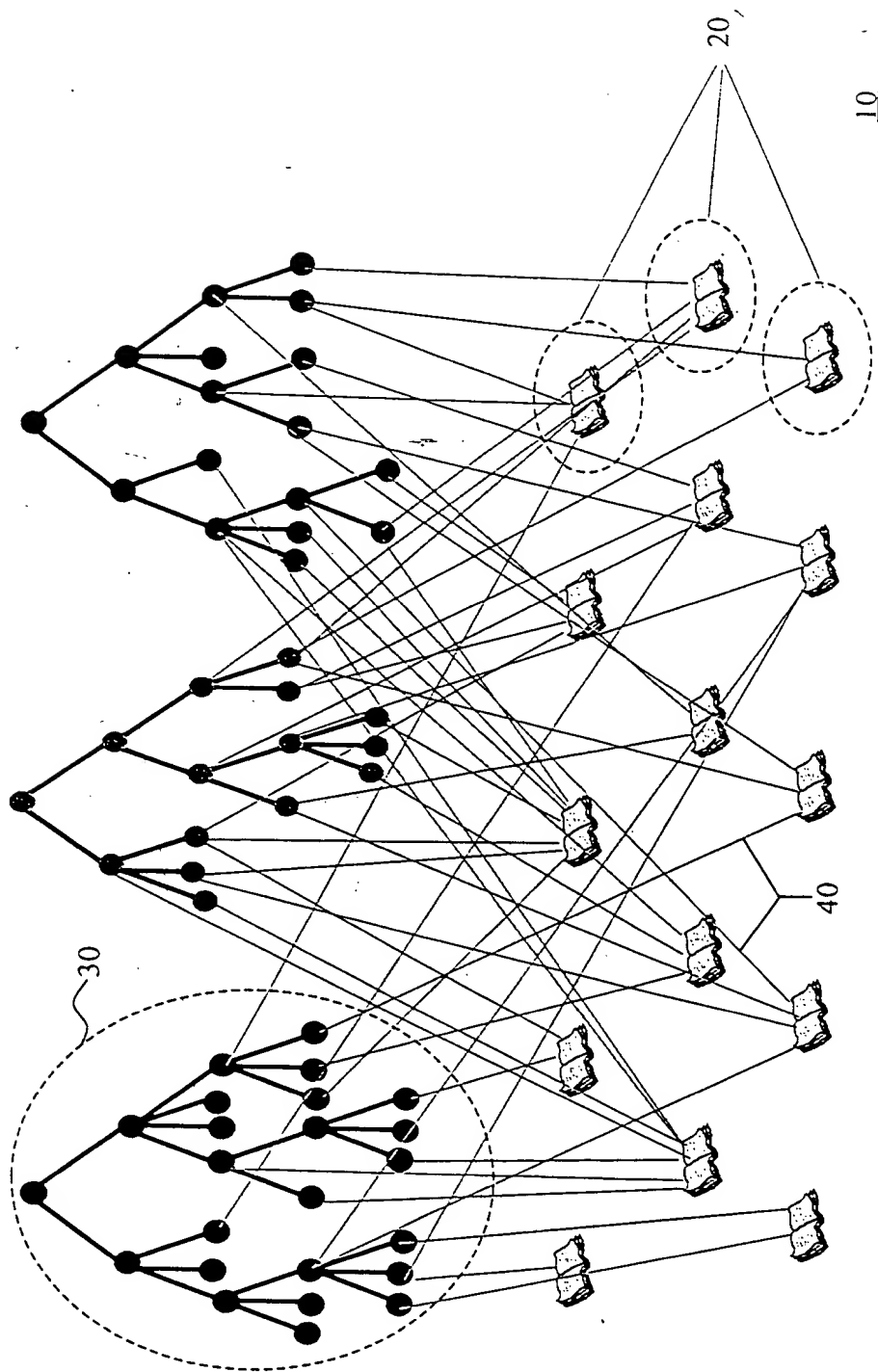


FIG. 1

A Knowledge Container


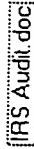

<i>Administrative meta-data</i>	<p>50</p> <p><author>Rev. Bill C. Wurtz</author> <creation date>6/7/89</creation date> <expiration date>12/31/99</expiration date></p>
<i>Taxonomy Tags</i>	<p>60</p> <p>Tax_Audit: 0.92 Tax_Evasion: 0.65 Fraud: 0.45</p>
<i>Marked Content</i>	<p>70</p> <p><P> In 1988, the <org>IRS</org> investigated <person>Scott Huffman</person>, looking for irregularities in income reporting. <P> ... <P>The <term>preliminary charges</term> included ...</p>
<i>Original Content</i>	<p>80</p> <p> </p>
<i>Links</i>	<p>90</p> <p> mugshot.bmp</p>

FIG. 2

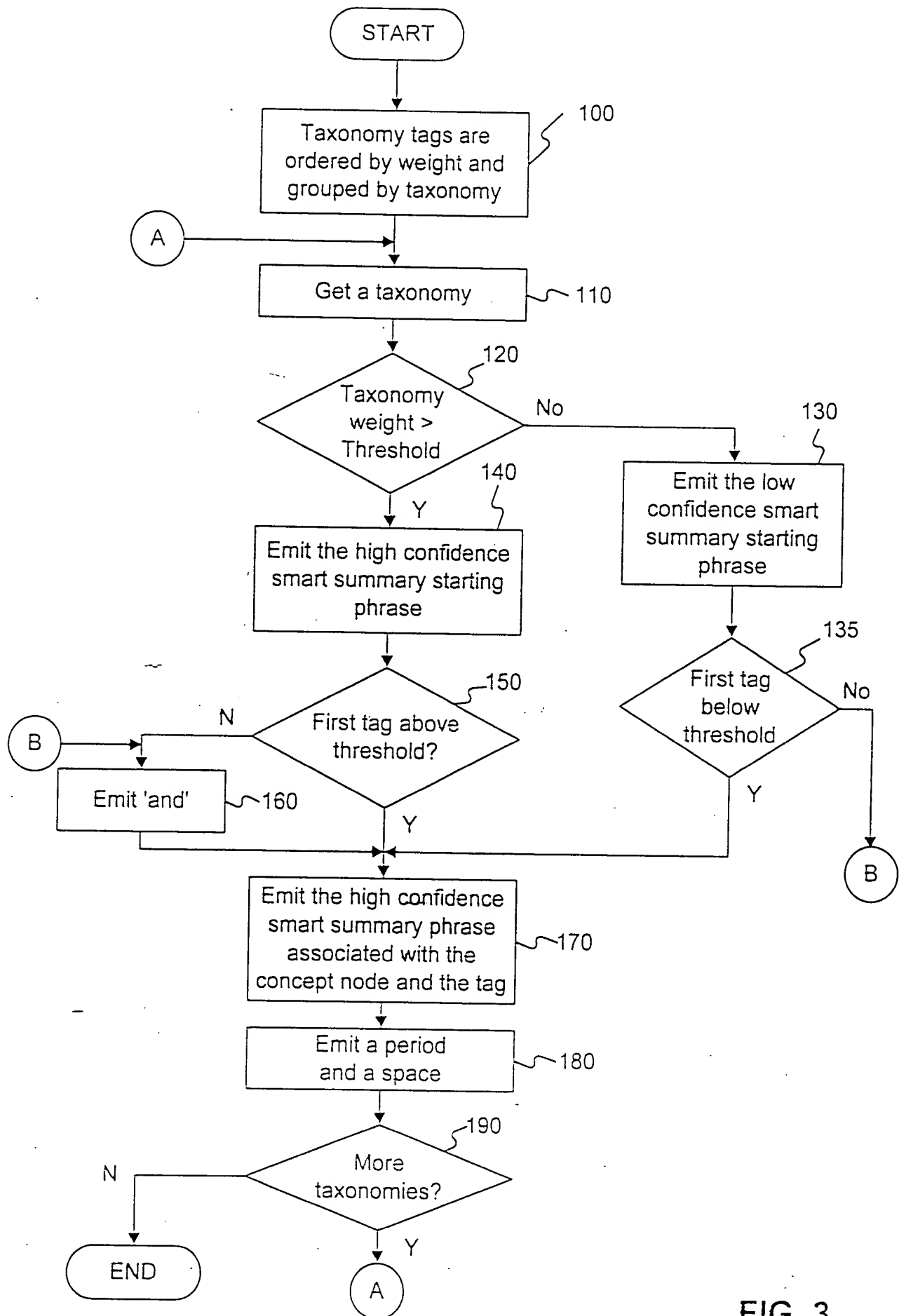


FIG. 3

Representative Taxonomy showing Types of Vehicles

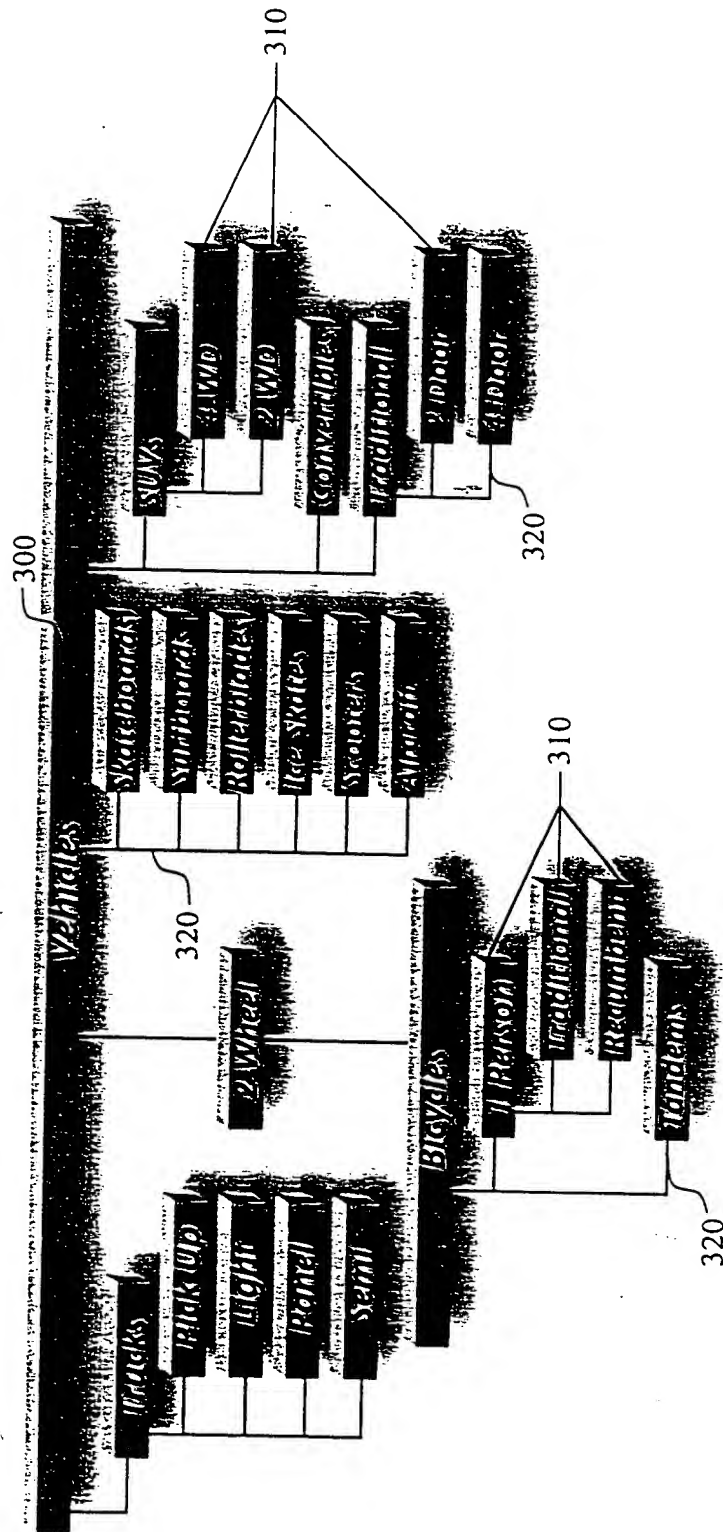


FIG. 4

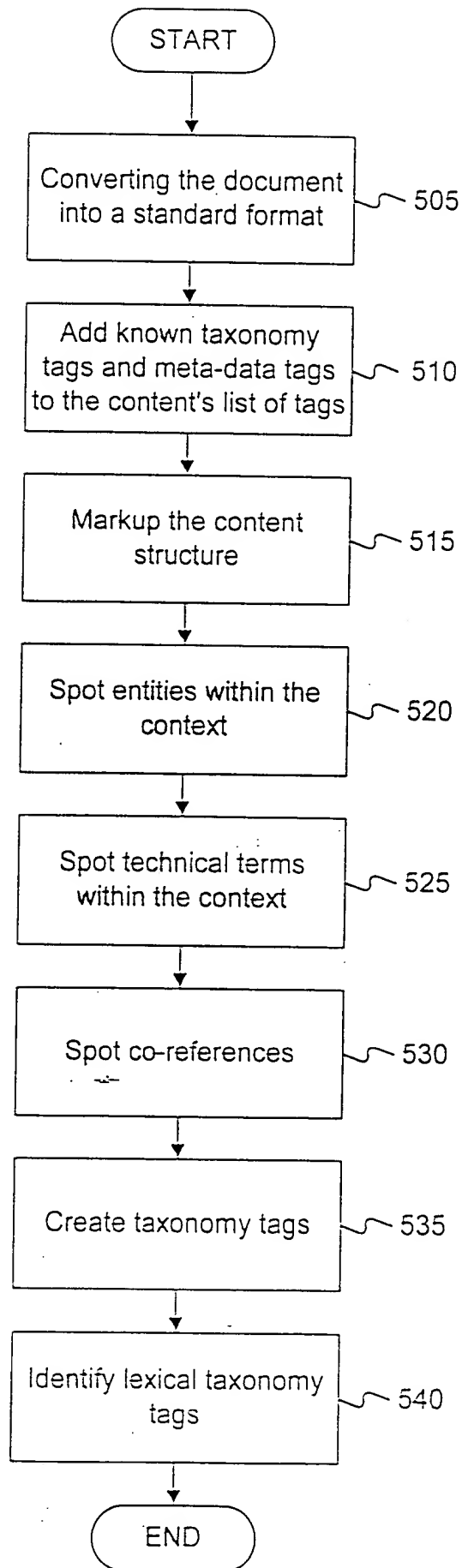
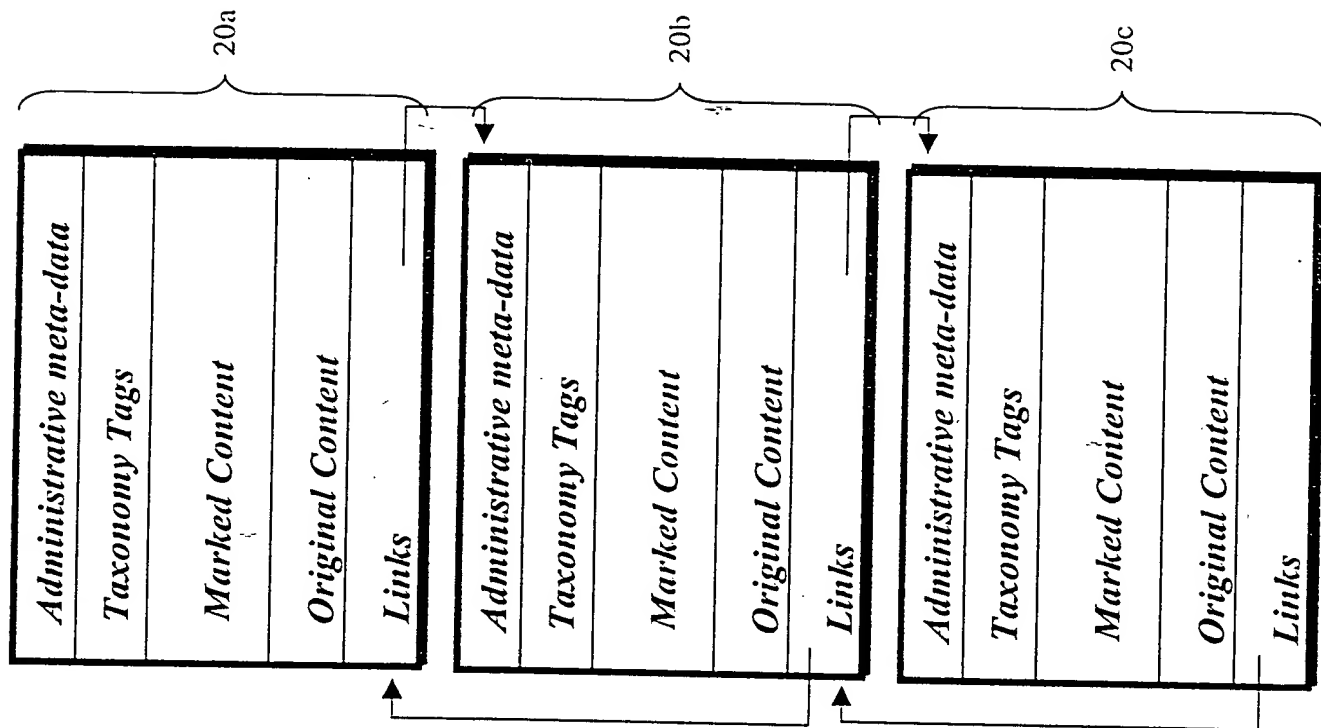


FIG. 5



Slice Representation

FIG. 6

Slicing (2)

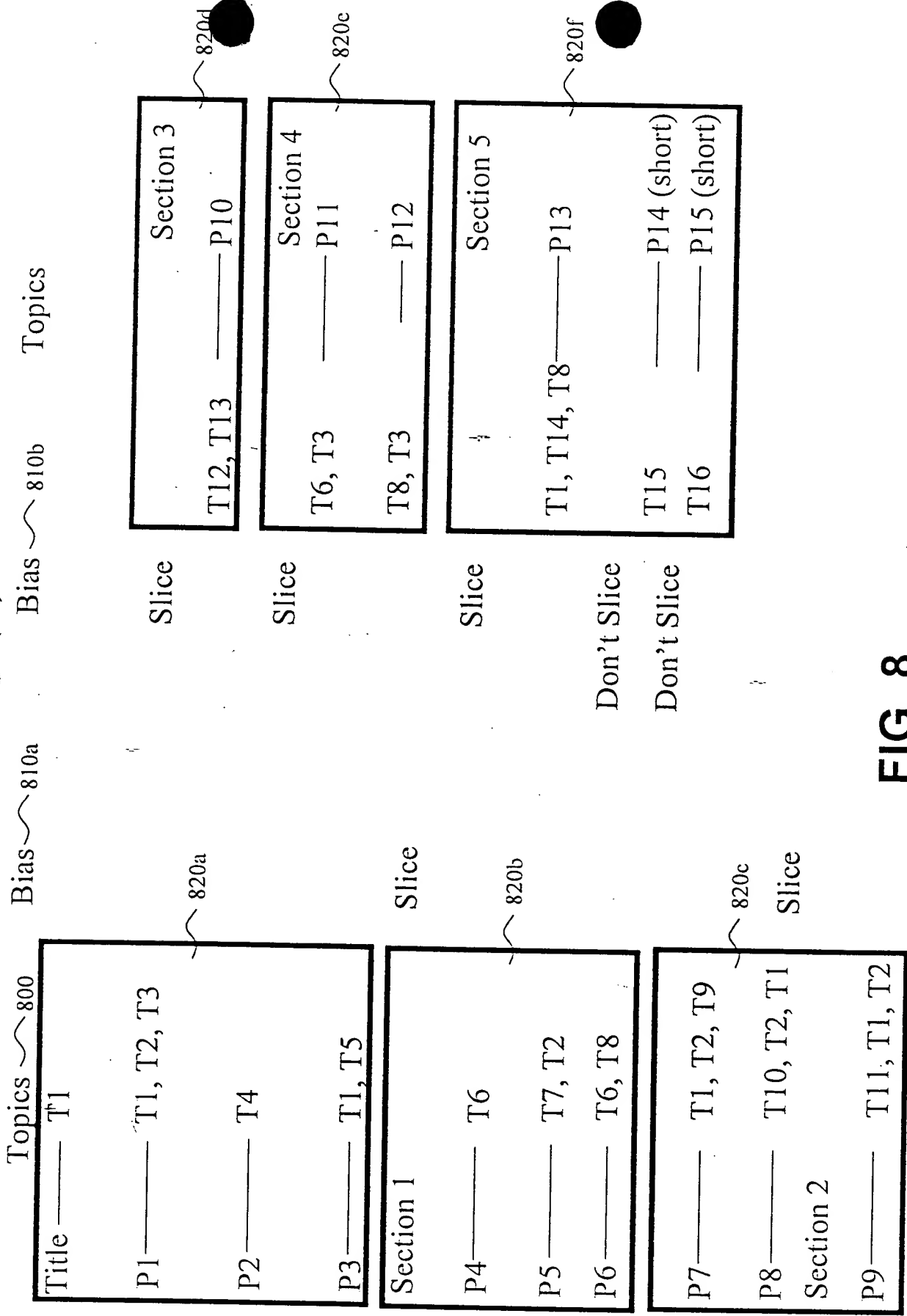


FIG. 8

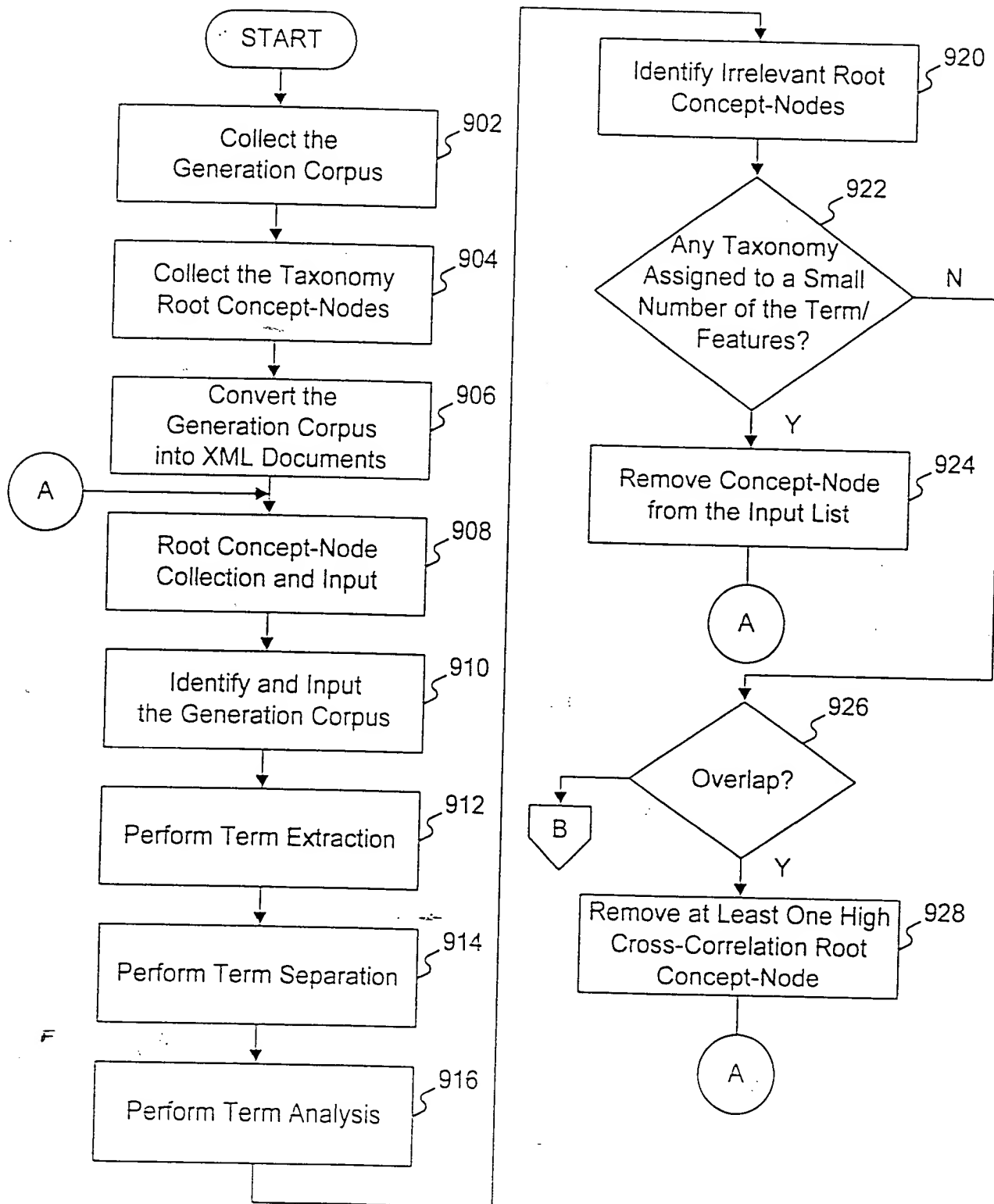


FIG. 9a

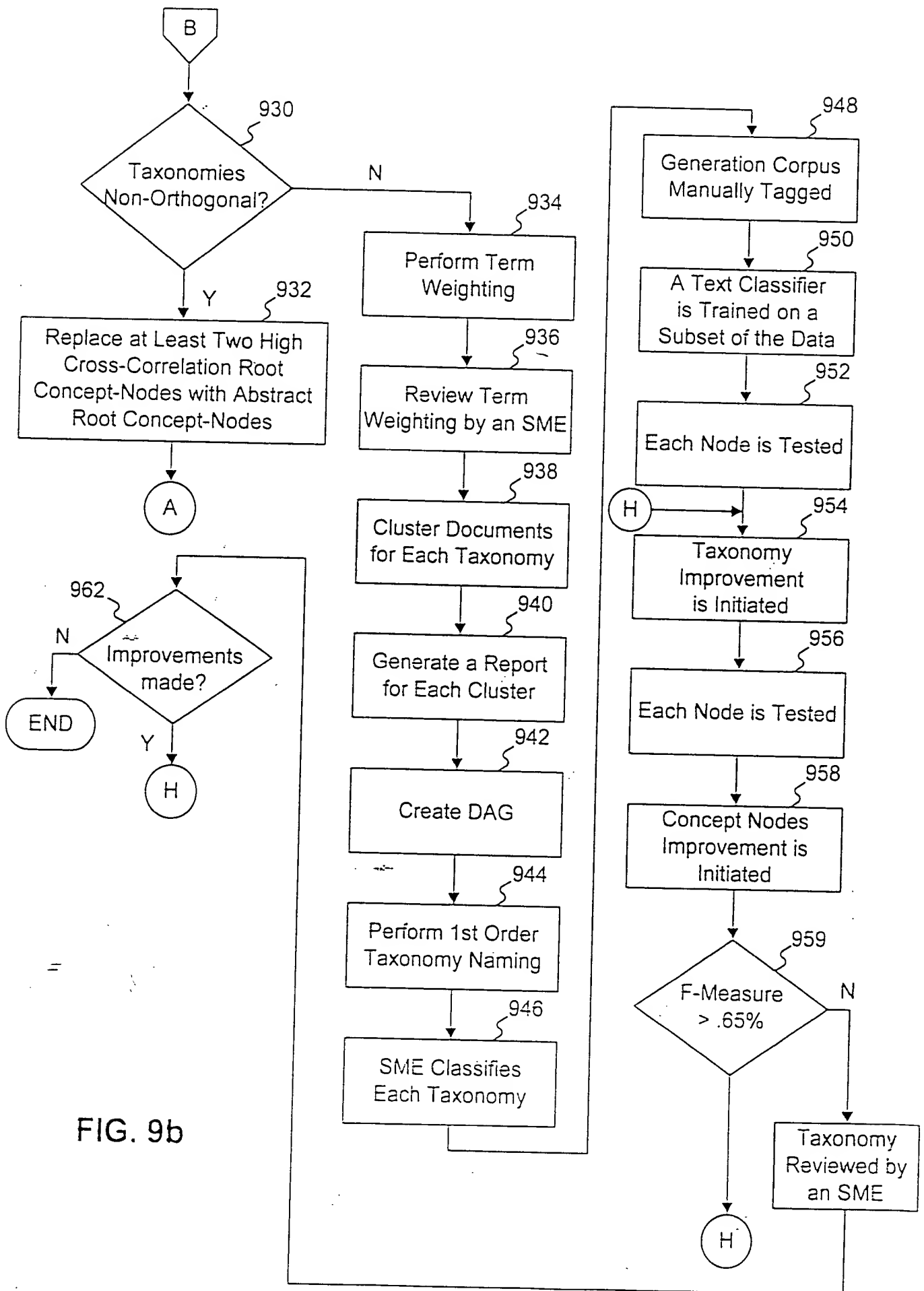


FIG. 9b

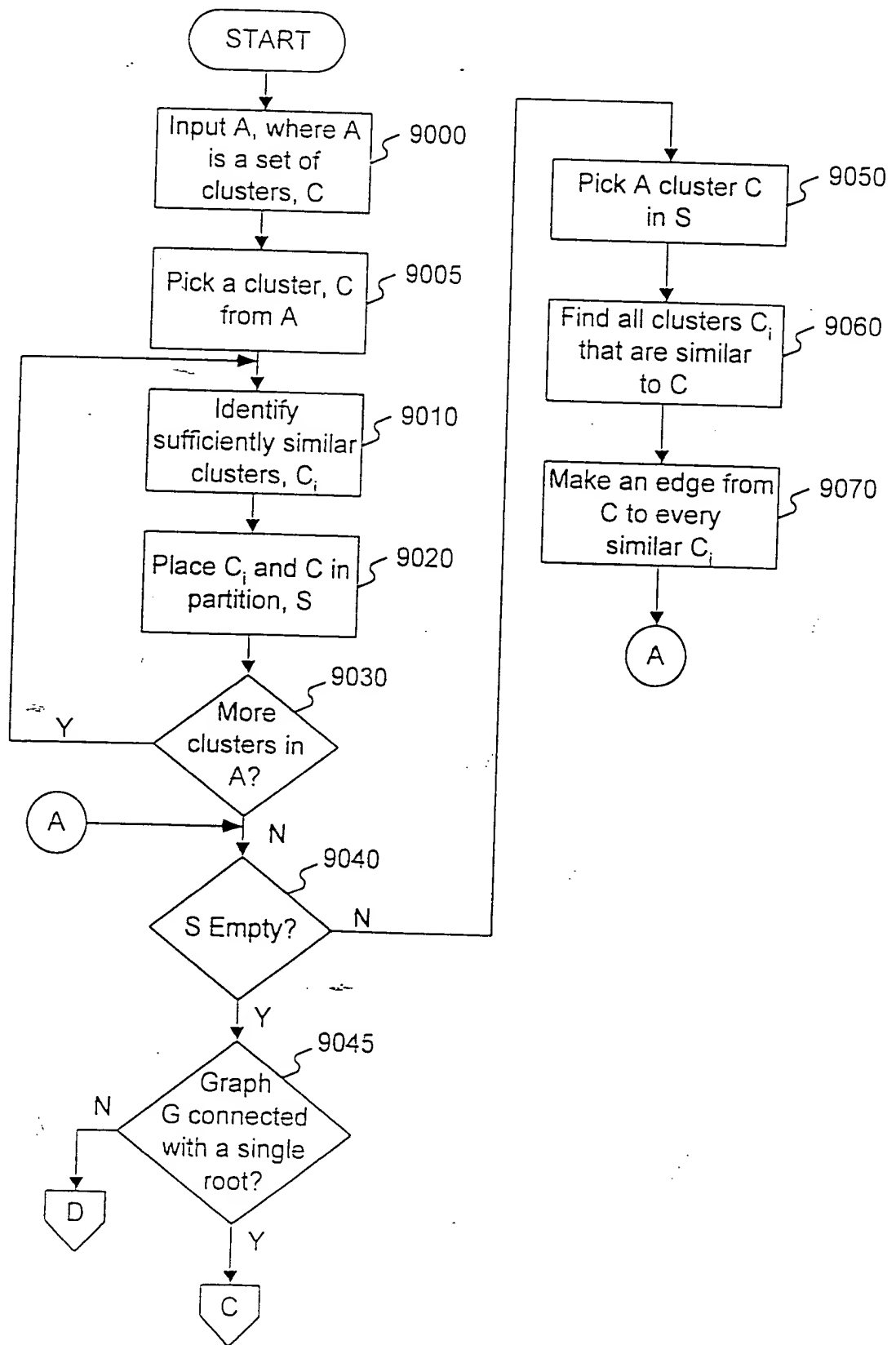


FIG. 9c

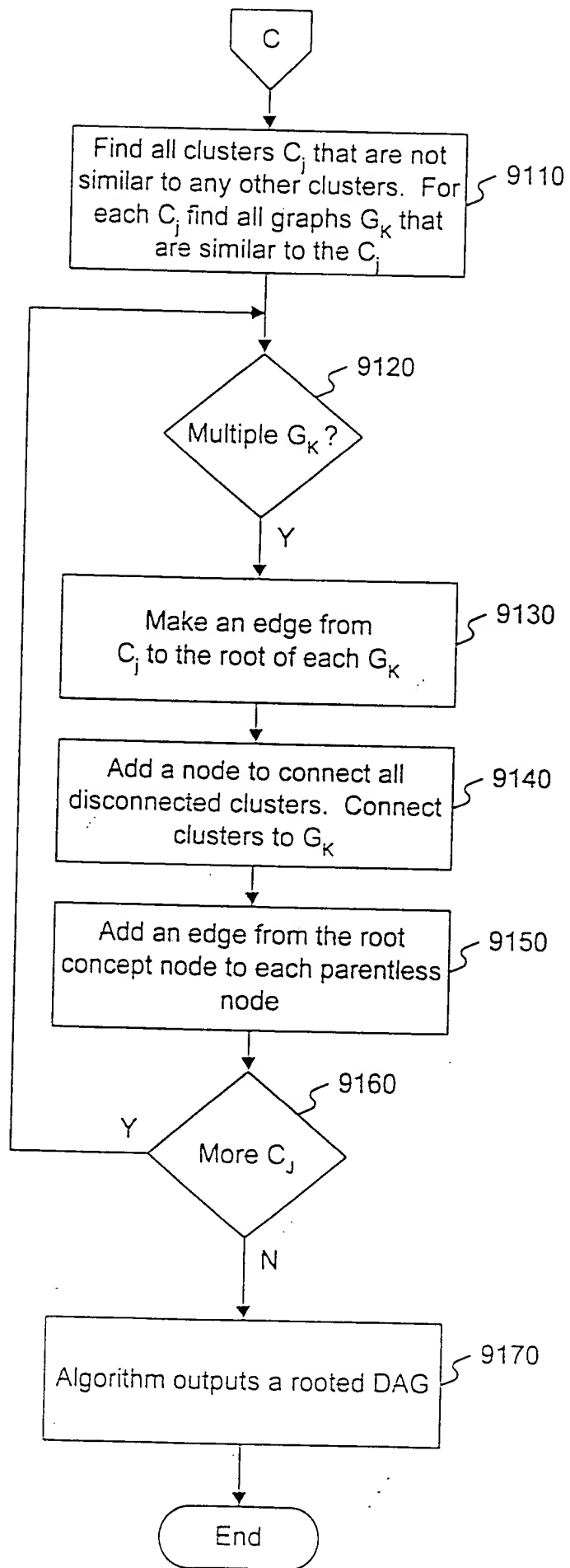
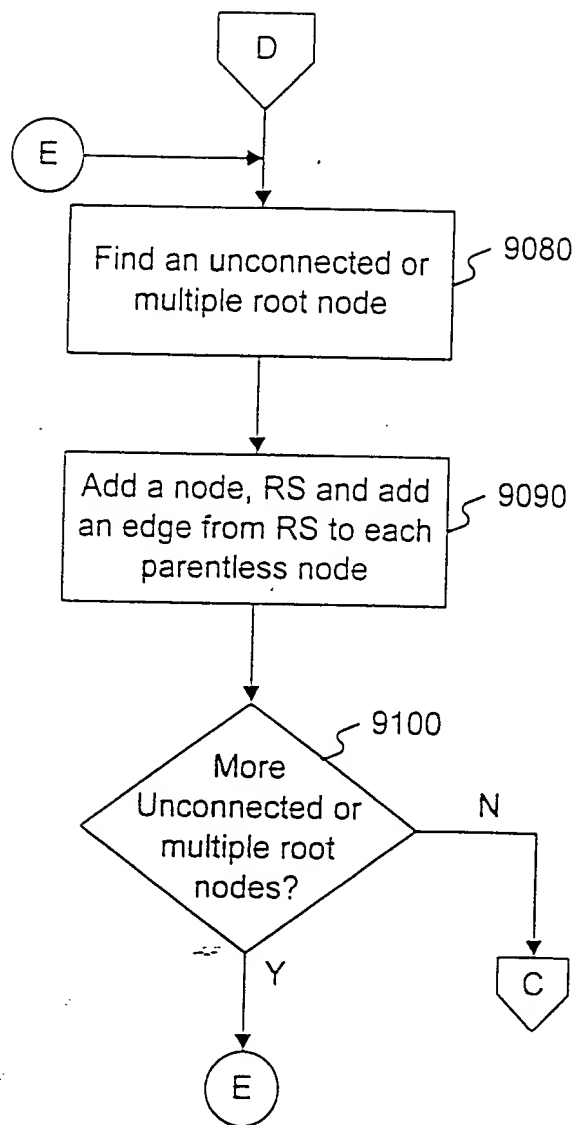


FIG. 9d

A Taxonomy of Document Sources

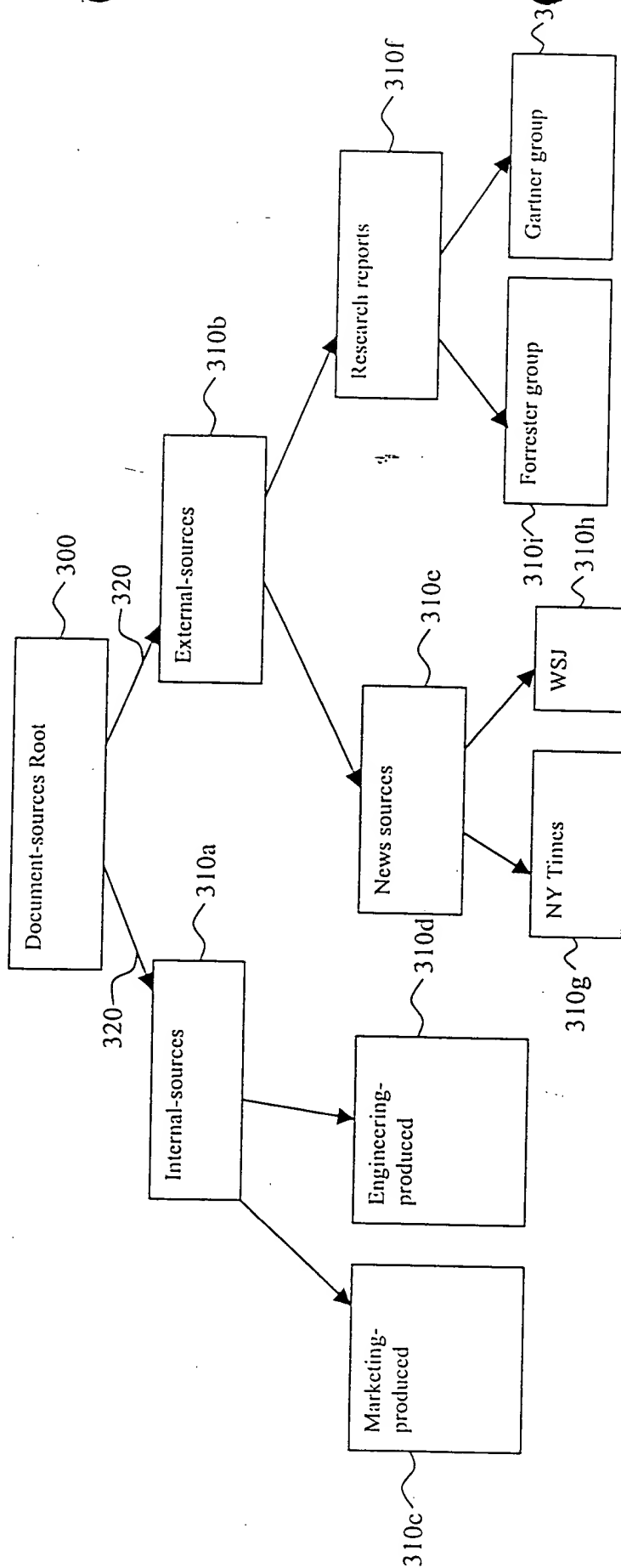
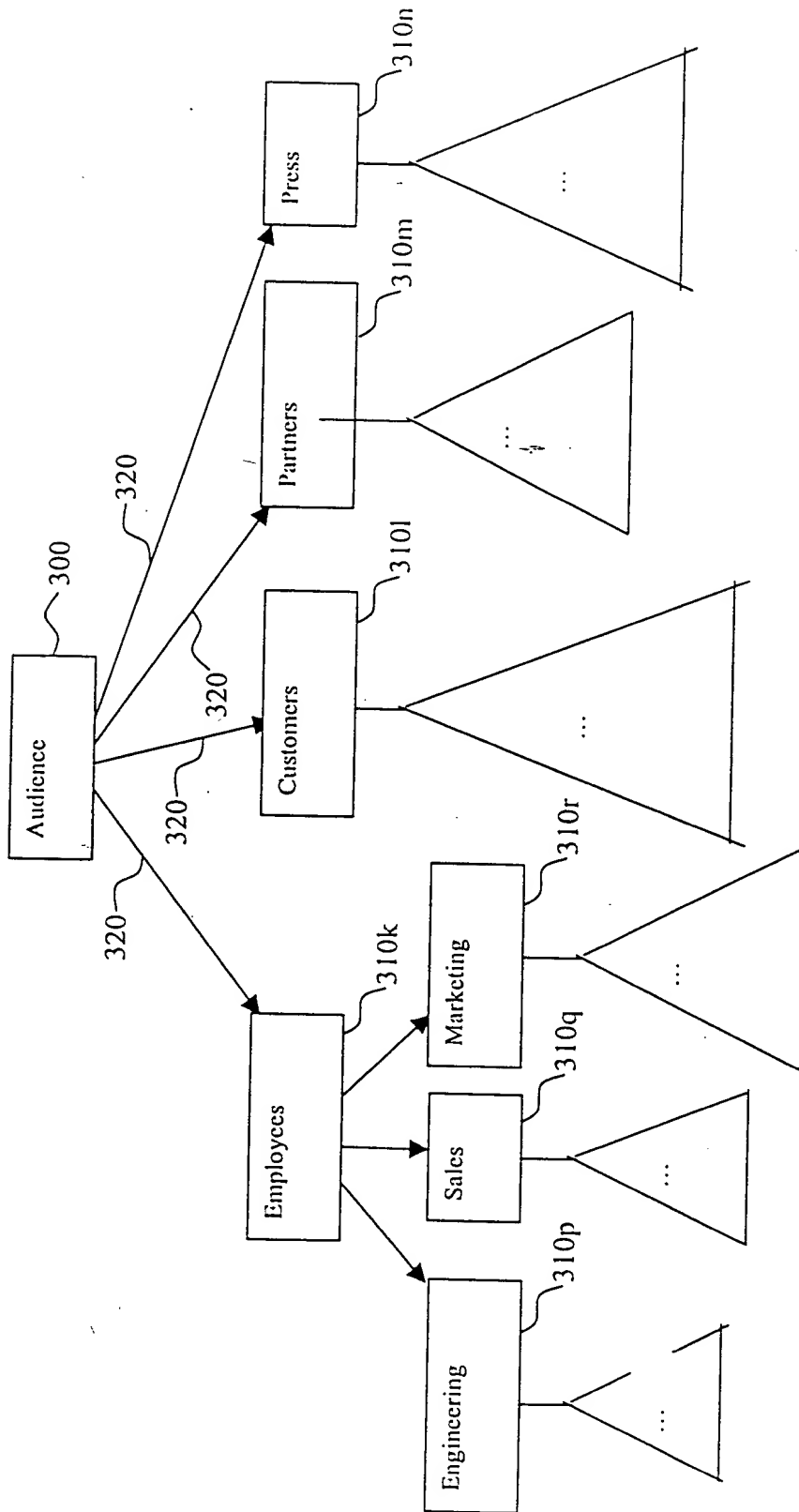


FIG. 10

Audience Taxonomy



30b

FIG. 11

Index Preparation---Step 1

For each node:

Make a "node document" which is the concatenation of all documents tagged to the node (here represented by ●).

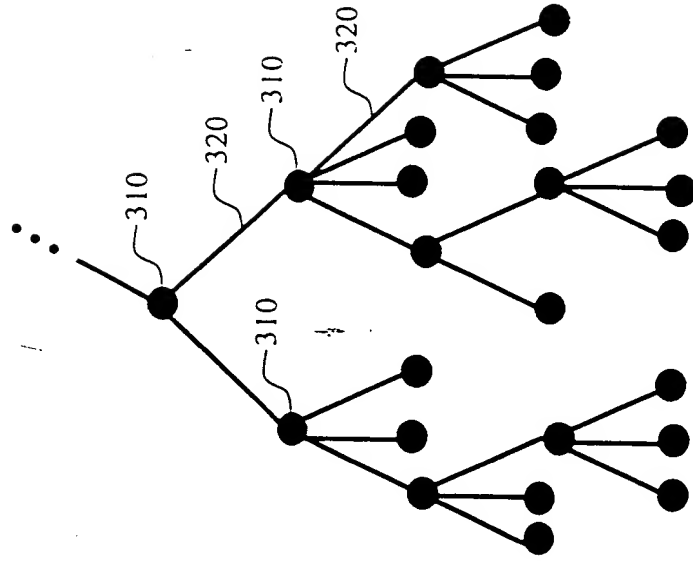






FIG. 12

Index Preparation---Step 2

Cluster the nodes according to similarity in vocabulary usage in the node-documents. Each lined-pattern represents a cluster.

-  - orange
-  - purple
-  - green
-  - blue

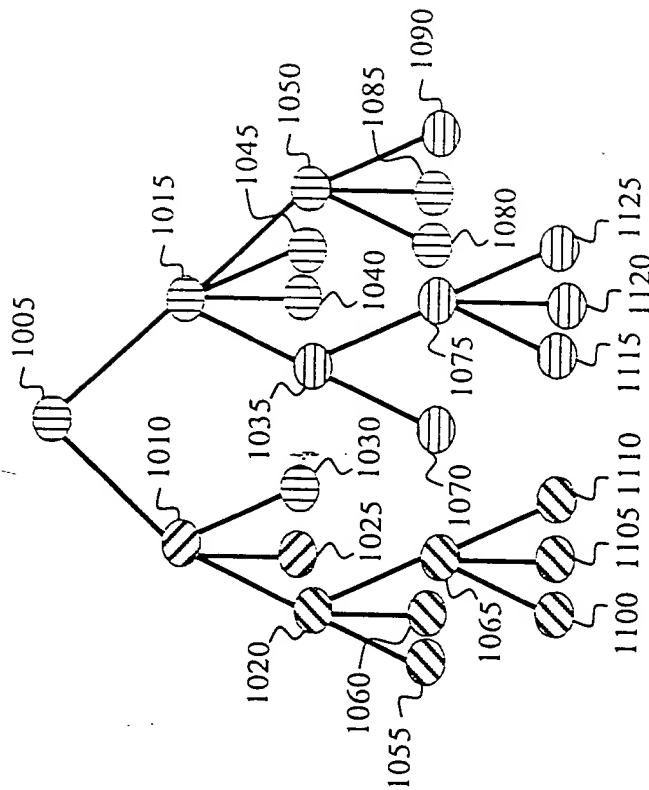
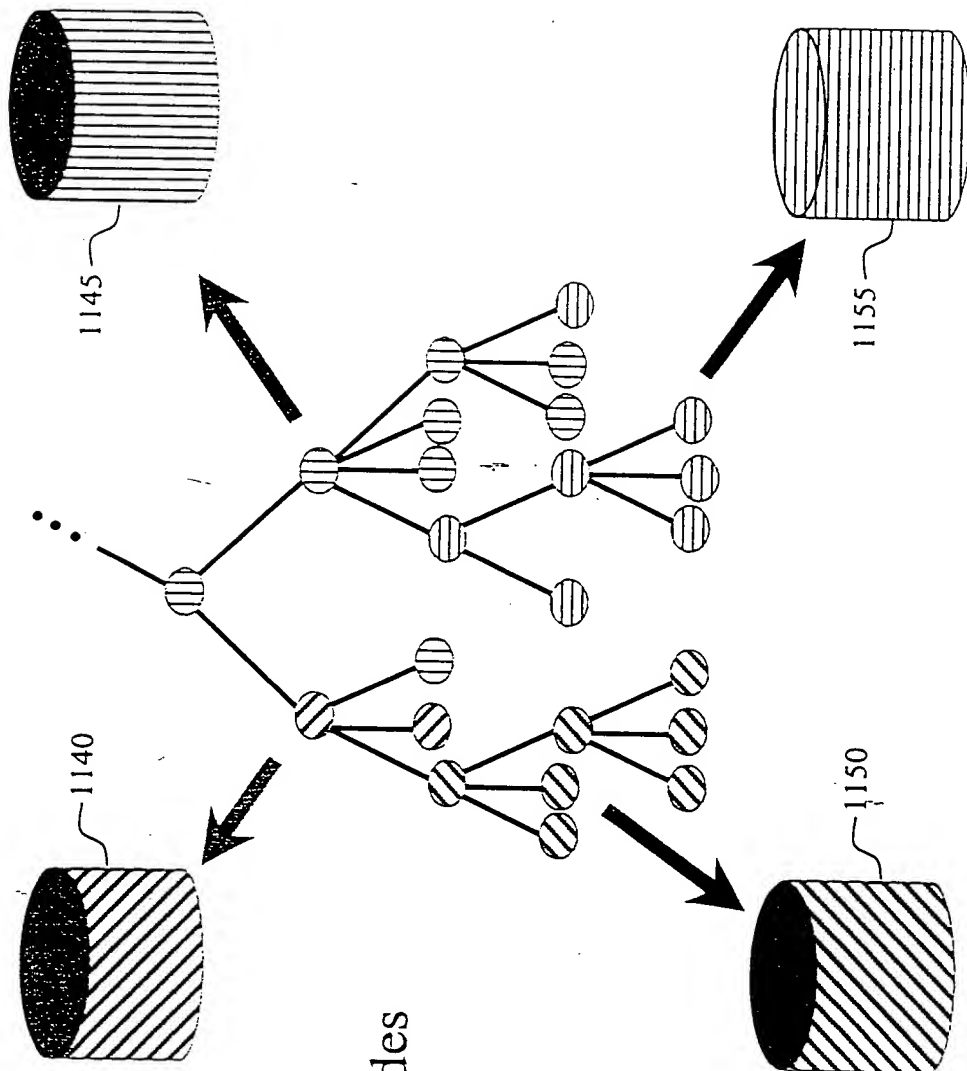


FIG. 13

Index Preparation---Step 3



For each cluster,
Build an index over the
documents tagged to nodes
in the cluster.





-  - orange
-  - purple
-  - green
-  - blue

FIG. 14

Region Designation: Marking

Identify nodes within distance D of a number N of query taxonomy tags. A node so identified is a *marked* node. Illustratively, here D is one and N is two.

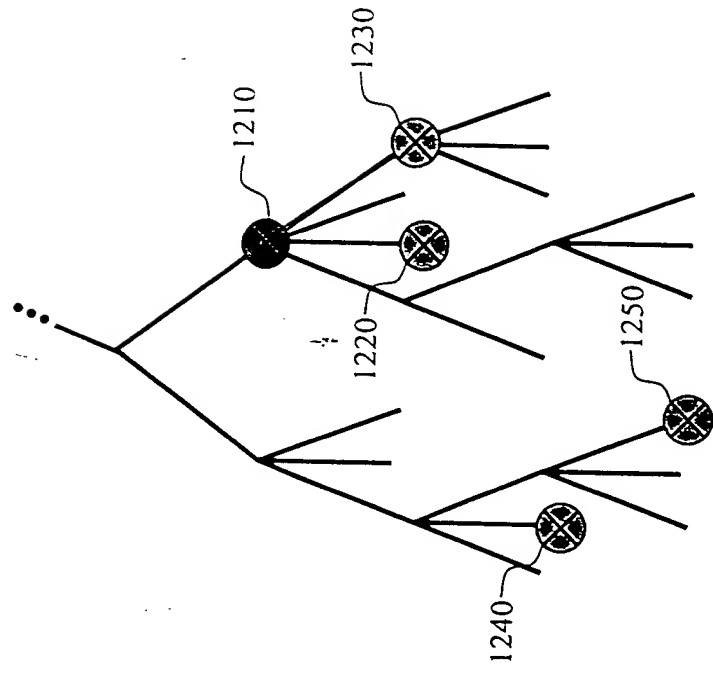


FIG. 15

Region Designation: Smoothing

Optionally, identify nodes within distance 1 of a query taxonomy tag or marked node. A node so identified is a *smoothed* node.

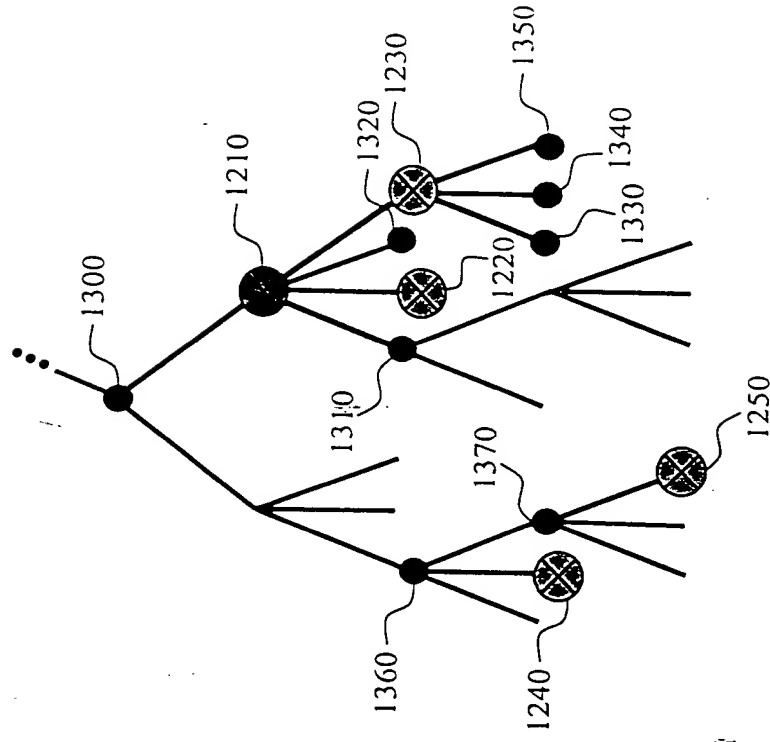


FIG. 16

Region Designation: Aggregation

Identify groups of query taxonomy tags, marked nodes, and smoothed nodes that, transitively, are within distance D' from each other. Illustratively, here D' is one.

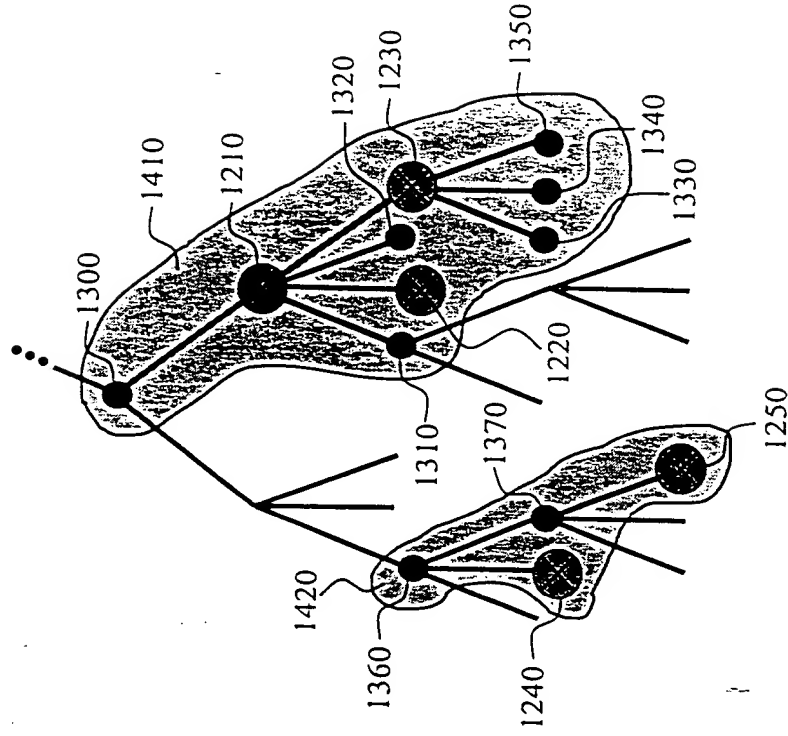




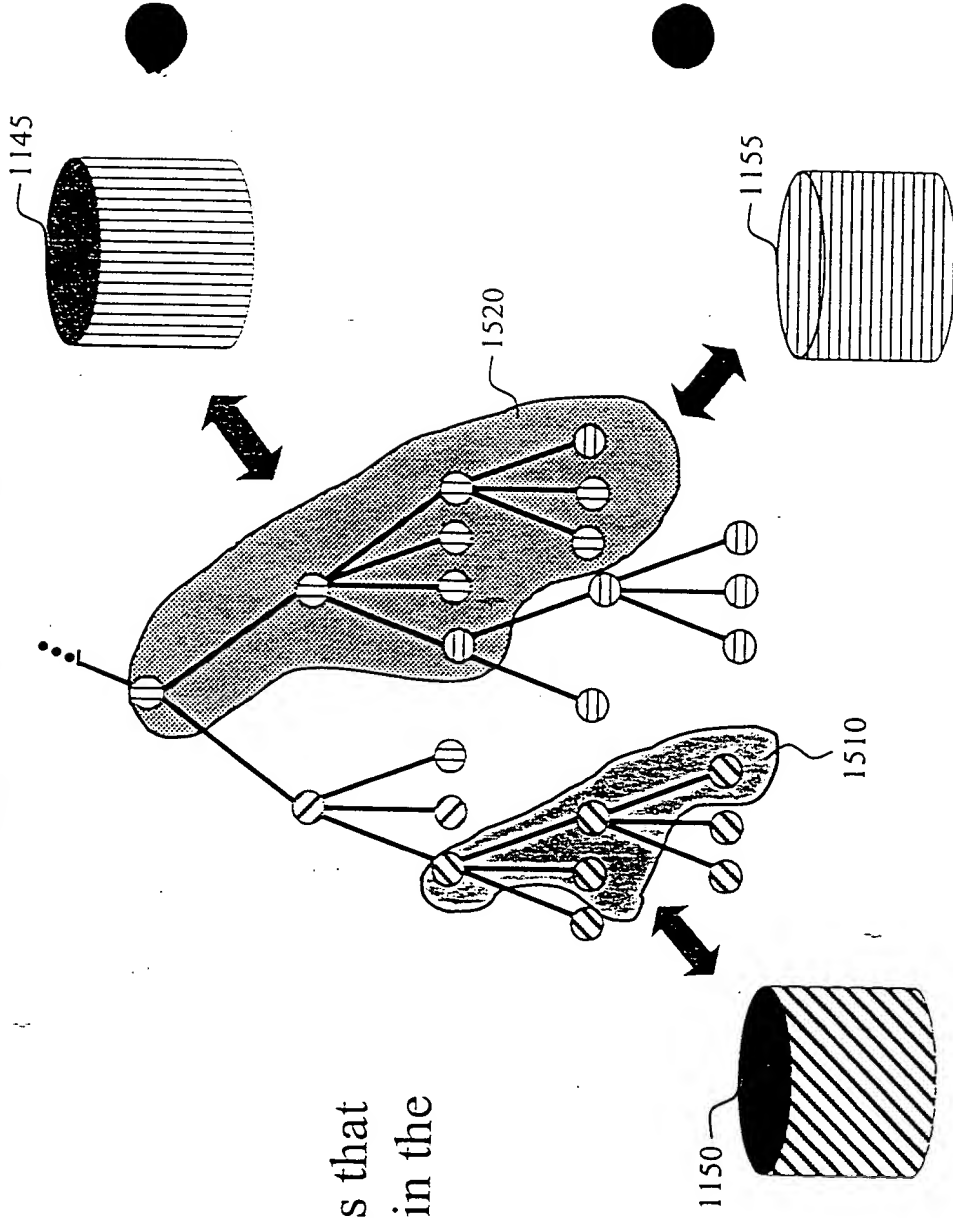


FIG. 17

Search: Index Identification

For each region,
Identify a set of indexes that
covers all of the nodes in the
region.

-  - orange
-  - purple
-  - green
-  - blue



1500

FIG. 18

Ranking: Search Engine Rank

For each document
the rank returned by
the search engine is
adjusted by ...

*For purposes of illustration,
rank is shown by distance
from the bottom of the picture.*

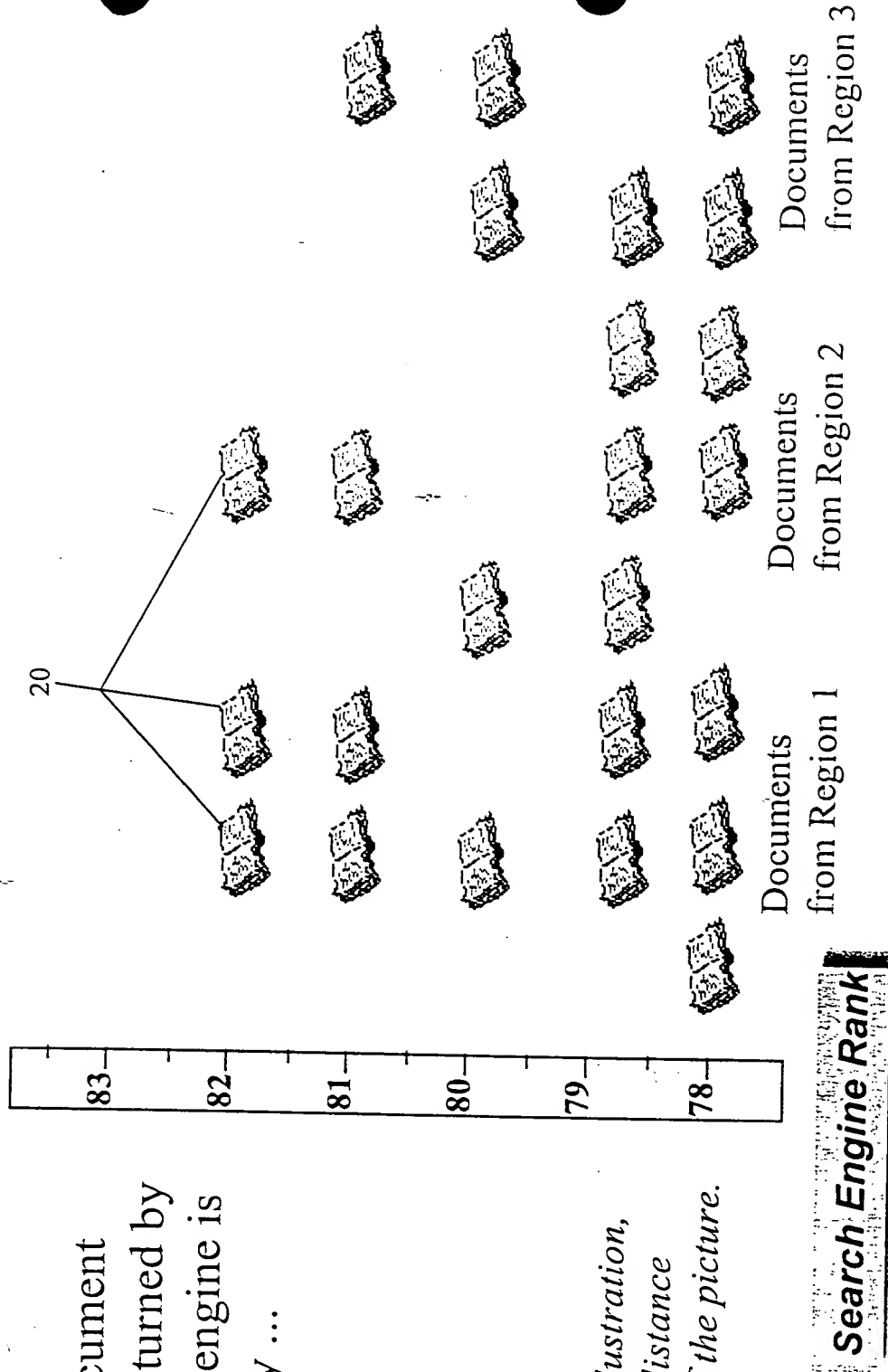
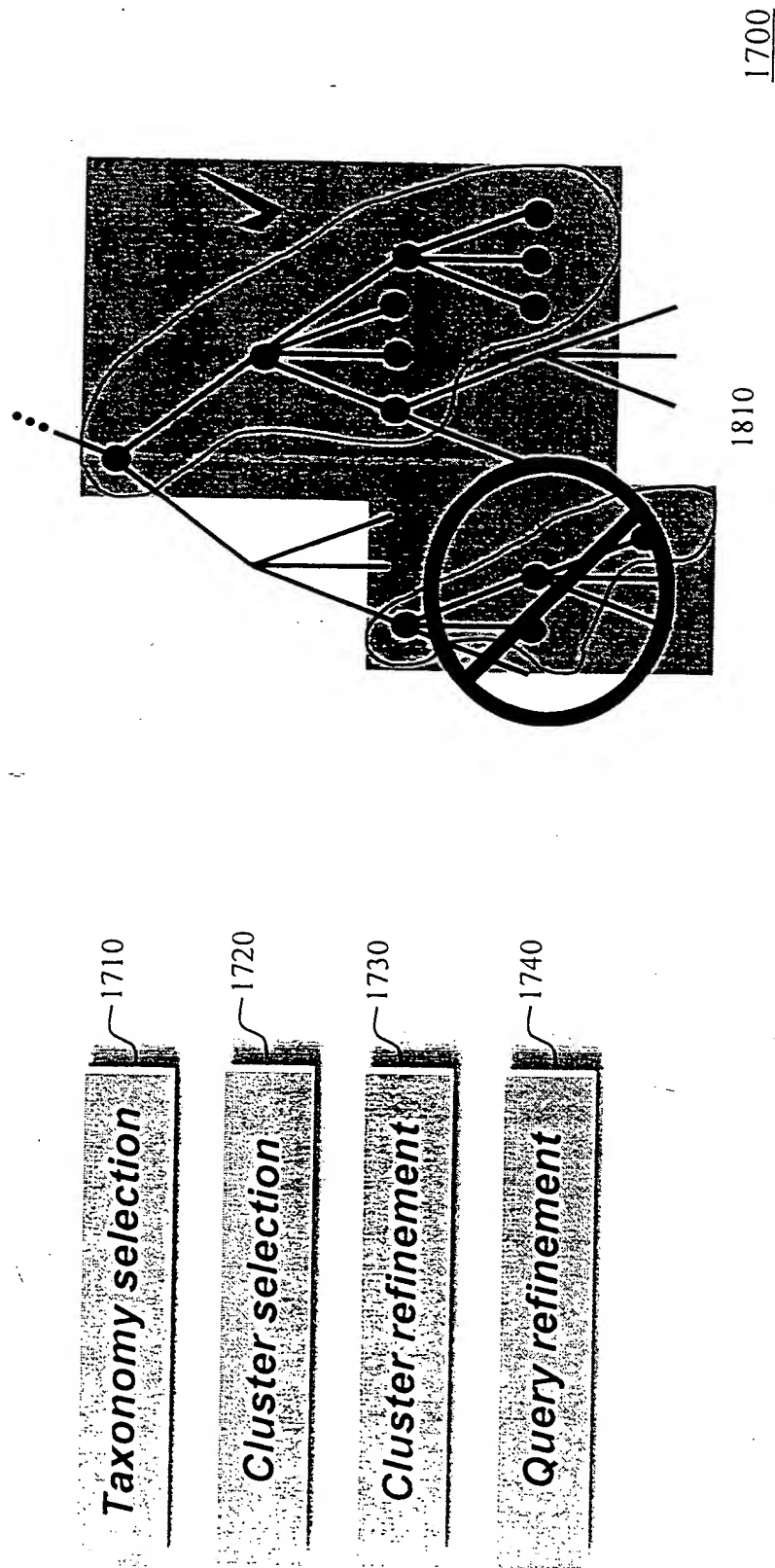


FIG. 19

Interactive Dialogue



The user can choose among the clusters.

FIG. 21

Topic ID	Topic name	# docs	# times topic was returned	# times topic was returned correctly	Precision	Recall	F-measure
v05	Securities	52	50	49	0.98	0.94	0.96
v11	Health Care	42	42	42	1.00	1.00	1.00
v07	Manufacturing	37	38	37	0.97	1.00	0.99
v12	Public Sector	35	35	35	1.00	1.00	1.00
v01	Financial Services Overview	30	29	29	1.00	0.97	0.98
v04	Insurance	30	30	30	1.00	1.00	1.00
v10	Utilities and Energy	28	27	27	1.00	0.96	0.98
v02	Banking (Business-to-Business)	27	27	27	1.00	1.00	1.00
v03	Banking (Business-to-Consumer)	26	26	26	1.00	1.00	1.00
v08	Retail	21	23	21	0.91	1.00	0.95
v09	Service Providers	12	12	12	1.00	1.00	1.00

* Summary:

conomy has 332 documents.
 documents on average have 1.02 tags
 call = 0.99
 precision = 0.99
 measure = 0.99

Per-topic statistics

Test On Train Report 1

[Edit](#) [View](#) [Go](#) [Refresh](#) [Home](#) [Back](#) [Forward](#) [Stop](#) [Reload](#) [Search](#) [Favourites](#) [History](#) [Channels](#) [Web](#) [FTP](#) [Mail](#)

C:\Cisco\Taxonomies\all\son\taxos\Cisco\Verticals\Variable Test\OnTrain\TrainDocResults.html

Per-document statistics Variable style

Document	human tags	topic spotter tags	# correct tags returned	Precision	Recall
12780	v05 (Securities)	v07 (Manufacturing)	0	0.00	0.00
12780	v01 (Financial Services Overview)	v05 (Securities)	0	0.00	0.00
12780	v05 (Securities)	v08 (Retail)	0	0.00	0.00
12780	v10 (Utilities and Energy)	v08 (Retail)	0	0.00	0.00
12780	v05 (Securities)	v05 (Securities)	1	1.00	1.00
12780	v11 (Health Care)	v11 (Health Care)	1	1.00	1.00
12780	v03 (Banking (Business-to-Consumer))	v03 (Banking (Business-to-Consumer))	1	1.00	1.00
12780	v01 (Financial Services)	v01 (Financial Services)	1	1.00	1.00

Test On Train Report 2

(v1.0)	Discriminating terms for v05 - Securities
2.42	Charles Schwab Securities Industry
1.77	Tax Relief Act
1.65	Securities and Exchange Commission
1.46	Merrill Lynch
1.42	Charles Schwab Corp.
1.41	compound annual growth rate
1.40	full-service firm
1.37	brokerage firm
1.37	minimum account balance
1.37	industry-wide testing
1.37	industrywide testing
1.33	discount brokerage
1.32	online brokerage
1.30	Ameritrade Holding Corporation
1.27	Wall Street
1.26	online broker
1.25	external resource
1.21	securities firm
1.19	IT staff
1.16	capacity expansion
1.16	technology development
1.10	implementation concern
1.09	Capacity Study
1.09	Securities Industry Association
1.08	mutual fund
1.08	infrastructure project
1.02	Tower Group
1.01	National Securities Clearing Corporation
00	staffing resource

op terms used by topic spotter:

1. Charles Schwab Securities Industry
2. Tax Relief Act
3. Securities and Exchange Commission
4. Merrill Lynch
5. Charles Schwab Corp.

Document	Classification	Human-assigned tags	Topic spotter tags	Topic spotter scores
13457		v05	v05 (Securities)	0.25
13460		v05	v05 (Securities)	0.35
13469		v05	v05 (Securities)	0.34
13502V055174650121798		v05	v05 (Securities)	0.20
13520		v05	v05 (Securities)	0.27
120541		v05	v05 (Securities)	0.18

FIG. 25

128 security	v05	v05 (Securities)	0.18
124 security	v05	v08 (Retail)	0.13
127 security	v05	v07 (Manufacturing)	0.25
125 security	v05	v05 (Securities)	0.28
126 security	v01 v03 v05	v03 (Banking (Business-to-Consumer)) v01 (Financial Services Overview)	0.25 0.23
126 security	v05	v05 (Securities)	0.21
127 security	v05	v05 (Securities)	0.21
128 security	v05	v05 (Securities)	0.24
129 security	v05	v05 (Securities)	0.30
130 security	v05	v05 (Securities)	0.32
131 security	v05	v05 (Securities)	0.19

Test On Train Report 5